May, 2014

System Safety Principles: A Multidisciplinary Engineering Perspective

Joseph H. Saleh, *Georgia Institute of Technology*
Karen B. Marais, *Purdue University*
Francesca M. Favaro, *Georgia Institute of Technology*

SYSTEM SAFETY PRINCIPLES: A MULTIDISCIPLINARY ENGINEERING PERSPECTIVE

Joseph H. Saleh [a, *], Karen B. Marais [b], Francesca M. Favaró [a]

[a] *School of Aerospace Engineering, Georgia Institute of Technology, USA*
[b] *School of Aeronautics and Astronautics, Purdue University, USA*

Abstract: System safety is of particular importance for many industries. Broadly speaking, it refers to the state or objective of striving to sustainably ensure accident prevention through actions on multiple safety levers (technical, organizational, and regulatory). While complementary to risk analysis, it is distinct in one important way: risk analysis is anticipatory rationality examining the possibility of adverse events (or accident scenarios), and the tools of risk analysis support and in some cases quantify various aspects of this analysis effort. The end-objective of risk analysis is to help identify and prioritize risks, inform risk management, and support risk communication. These tools however do not provide design or operational guidelines and principles for eliminating or mitigating risks. Such considerations fall within the purview of system safety.

In this work, we propose a set of five safety principles, which are domain-independent, technologically agnostic, and broadly applicable across industries. While there is a proliferation of detailed safety measures (tactics) in specific areas and industries, a synthesis of high-level safety principles or strategies that are independent of any particular instantiation, and from which specific safety measures can be derived or related to, has pedagogical value and fulfills an important role in safety training and education. Such synthesis effort also supports creativity and technical ingenuity in the workforce for deriving specific safety measures, and for implementing these principles and handling specific local or new risks. Our set of safety principles includes: (1) the fail-safe principle; (2) the safety margins principle; (3) the un-graduated response principle (under which we subsume the traditional "inherently safe design" principle); (4) the defense-in-depth principle; and (5) the observability-in-depth principle. We carefully examine each principle and provide examples that illustrate their use and implementation. We relate these principles to the notions of hazard level, accident sequence, and conditional probabilities of further hazard escalation or advancement of an accident sequence. These principles are a useful addition to the intellectual toolkit of engineers, decision-makers, and anyone interested in safety issues, and they provide helpful guidelines during system design and risk management efforts.

Keywords: Safety principles; fail-safe; safety margins; defense-in-depth; observability-in-depth; system safety.

[*] Corresponding author. Tel: + 1 404 385 6711; fax: + 1 404 894 2760
E-mail address: jsaleh@gatech.edu (J. H. Saleh)

1. Introduction

In this work, we provide a synthesis of system safety principles, and we examine their use and implementation in different settings. These high-level principles are domain-independent, technologically agnostic, and broadly applicable across various industries. The objective of this synthesis is mainly educational, and it is meant to serve a useful role in safety training and education. It can also support creativity and technical ingenuity in the workforce to conceive and implement these principles in new or different ways to handle specific local hazards, or new and emerging ones.

System safety is particularly important for many industries, such as the nuclear and the airline industries, and broadly speaking, it refers to the state or objective of striving to sustainably ensure accident prevention through actions on multiple safety levers, be they technical, organizational, or regulatory.

Detailed safety measures abound for dealing with particular hazards, such as electrocution and fire, for example. But the proliferation of safety measures in domain-specific areas is not conducive to adapting or devising safety measures to handle new or emerging hazards, and more importantly it is not well suited for general safety education and training of engineers and decision-makers. What is more useful for such audiences are general safety principles and strategies, from which specific safety measures can be derived or related to. The distinction between specific safety measures and general safety principles is somewhat similar to that between *tactics* and *strategy* in a military context: the former relates to specific moves and dispositions to achieve a local objective (e.g., moving soldiers and equipment, engaging in a skirmish or a battle), whereas the latter, *strategy*, relates to broader considerations for planning and organizing to succeed in a general conflict (e.g., war) with an opponent. More details on this distinction along with several examples follow in the subsequent sections.

Considerations of system safety, and the related safety principles, while complementary to risk analysis, are distinct in one important way: risk analysis is anticipatory rationality examining the possibility of adverse events or accident scenarios, and the tools of risk analysis support, and in some cases help quantify various aspects of this analysis effort. Risk analysis has been described as addressing three main questions (Apostolakis, 2004; Kaplan and Garrick, 1981):

(1) What can go wrong?
(2) How likely it is?
(3) What would be the consequences?

The end-objective of risk analysis is to help identify and prioritize risks, inform risk management, and support risk communication. These tools however do not provide design or operational guidelines or principles for eliminating or mitigating risks, and they are mainly concerned with process[1]. Such considerations fall within the purview of system safety. The safety principles examined in this work provide guidelines and conceptual support during system design and operation for addressing the most important follow-up question, namely:

(4) What are you going to do about it [what can go wrong]? Or how are you going to defend against it?

Previous efforts at synthesizing safety principles include the works by Haddon (1980a, 1980b), Möller and Hansson (2008), Kletz (1978; 1998, and subsequent works), and Khan and Amyotte (2003). The present article follows in the spirit of these works, and in some cases it builds and expands on them. These works are briefly reviewed in Section 2. Section 3 presents and examines the proposed set of safety principles. Section 4 concludes this work.

2. Brief literature review of safety principles

2.1 Haddon's safety principles and their energy-centric underpinning

Haddon's work (1980a, 1980b) is a landmark in the study of the epidemiology of injury and accident prevention[2]. It is grounded in the public health realm and conceptualizes injury as an epidemiologic problem with agent(s), hosts, vectors (for the transmission of injury-producing elements), and the environment (physical and social). Haddon's contributions build on previous work by Gibson (1964; first presented in 1961) in which the agents of injury were first identified as various forms of energy—this idea is referred to nowadays as the energy model of accidents (Saleh *et* al., 2010):

> "*Man [...] responds to the flux of energies which surround him—[...] mechanical, thermal, and chemical. Some limited fields and ranges of energy produce stimuli*

[1] They can help assess the effectiveness of a particular implementation of a safety principle once it has been devised. The literature on risk analysis is extensive (the topic is not the focus of the present work). The reader interested in a good introduction to risk analysis and management may consult the excellent works by Kaplan and Garrick (1981), Pate-Cornell (1997), Rasmussen (1997) and the ISO 31000 (2009) and ISO 31010 (2009) standards.

[2] Rivara et al. (2001) consider an earlier work by Haddon *et* al. published in 1964 and entitled *Accident Research: Methods and Approaches* as "one of the most important milestones in the development of injury research" worldwide.

> *for the sense organs; others induce physiological adjustments; still others produce injuries. […] Injuries to a living organism can be produced only by some energy interchange*." (Gibson, 1964)

Haddon expanded on this energy basis of injuries, and the safety strategies he devised are fundamentally tied to this perspective:

> "*A major class of [adverse] phenomena involves the transfer of energy in such ways and amounts, and at such rapid rates that inanimate and animate structures are damaged. The harmful interactions with people and properties of […] projectiles, moving vehicles, ionizing radiation, conflagrations […] illustrate this class of phenomena*." (Haddon, 1980a; quote from earlier work by Haddon)

Haddon's development of the energy model led him to propose a set of safety strategies to guide the development of injury control mechanisms and safety interventions. The distinction between a safety *strategy* and a safety *tactic*/measure, previously noted, is important to keep in mind, and to be able to appreciate the distinctive contribution of Haddon. A safety strategy can be implemented in a variety of ways and measures, and domain-specific knowledge is required, e.g., design and operation of a splitter tower at a refinery, as well as creativity and technical ingenuity to translate a safety principle into a specific safety measure (examples are provided hereafter and further discussed in Section 3). Haddon's safety strategies include the following:

i. Reduce the amount of hazard/energy brought into being in the first place (e.g., reduce speed of vehicles in the context of traffic safety);

ii. Modify or reduce the rate of release of hazard/energy from its source (e.g., shutoff valves, nuclear reactor control rod);

iii. Separate in time and space the energy source (hazard) from that which is to be protected; eliminate the intersections of hazard/energy and susceptible structure or individuals (Haddon argues that the use of sidewalks and phasing of pedestrian and vehicle traffic is one example of the implementation of this strategy; other examples include the more common use of physical barriers to separate hazard sources from individuals). This principle was described as preventing the etiological agent, the energy source, from reaching the susceptible host;

iv. Make what is to be protected more resistant to damage from the hazard/energy (e.g., make structures more fire- and earthquake-resistant[3]; Runyam (2003) in discussing Haddon's principles provide the example of a bullet-proof garment

[3] How to do this would be an example of a specific safety measure.

as an example of the implementation of this principle for dealing with injuries from handguns).

A detailed discussion of these principles can be found in Haddon (1980a, 1980b). These principles remain according to Runyan (2003) “an excellent brainstorming tool for developing ideas about a range” of possible safety interventions. Haddon’s principles can be found in Section 3, subsumed in part under the *un-graduated response* and the *defense-in-depth* principles, although expressed differently and tailored toward system accidents.

2.2 Möller and Hansson synthesis of safety principles and the reduction of risk and uncertainty

Möller and Hansson (2008) provided a much needed recent synthesis of engineering safety principles. The authors recognized that despite the importance of the topic, “there is a lack of general accounts of safety principles [in] the literature. The treatment is normally piecemeal, focusing only on specific [safety measures]” or methodological issues in probabilistic risk analysis (these topics are important, but they are downstream of the concerns with safety principles, as mentioned previously).

The authors provided a list of 24 safety principles and subsumed them under four broad categories: (1) inherently safe design; (2) safety reserves; (3) safe fail; and (4) procedural safeguards. Although the authors did not acknowledge an energy basis of accidents, they related their safety principles to an important aspect, namely the probabilistic consequences of the implementation of said principles. To this effect, the authors argued, and provided ample examples, that the end-objective of any safety principle is “not only to reduce the probabilities of negative events that have been foreseen and for which probability estimates have been provided, but also [to reduce] epistemic uncertainty.”

The distinction between Möller and Hansson’s view and Haddon’s is worth highlighting: in recognizing the energy basis of accidents, Haddon identified safety principles that are meant to limit or contain uncontrolled releases of energy, and to segregate in time and space energy sources from that which is to be protected. Möller and Hansson (2008) synthesized their safety principles under the various means by which they reduce risk and/or uncertainty. Both views are important and complementary, and while the former is content-centric and allows some creativity in deriving novel safety measures (safety principles relate to the handling of energy sources and releases), the latter highlights process-centric issues and ways for evaluating safety measures (safety principles relate to ways for reducing risks and/or uncertainties about accident occurrence).

One limitation in Möller and Hansson's work is that several themes described as safety principles are not principles, but specific safety measures in some cases, and too vague categories to be meaningful principles in other cases. For example, the authors list timed replacement, procedural safeguards, and redundancy as safety principles. They are not: *timed replacement* for instance is one maintenance technique (among many others); *procedural safeguards* is a broad descriptive term that is difficult to translate into an actual safety principle; and redundancy[4] is more akin to a specific reliability-improvement tactic, but it can backfire through common-cause failure (Hoepfer et al., 2009). As such, redundancy, while an important consideration for engineers, cannot be taken as an unquestionable or always-dependable safety measure. We revisit some of these considerations in more detail in Section 3.

2.3. Kletz's inherent safety design principle and its pillars

Kletz first outlined the basis of the inherent safety design principle in 1978 in an article titled, "What you don't have, can't leak". The work was based on the author's experience with the chemical industry, and the principle was later further extended by the author (Kletz, 1998), as well as by others, for example Kletz and Amyotte (2010) and Khan and Amyotte (2003).

The motivation for this principle came from a simple observation:

> *"If we could design our plants so that they use safer raw materials and intermediates, or not so much hazardous ones, or use hazardous ones at lower temperatures and pressures, then we would avoid, rather [than have to] solve our [safety] problems. Such plants can be described as [inherently] safe."* (Kletz, 1978)

Kletz later formulated the inherent safety principle succinctly as one that guides the development of inherently safer designs:

> *"An inherently safer design is one that avoids hazards instead of controlling them, particularly by removing or reducing the amount of hazardous material in the plant or the number of hazardous operations. [...] The words "inherently safer" imply that the plant or operation is safer because of its very nature, and not because [protective] equipment has been added on to make it safer." (*Kletz, 1998)

[4] "Redundancy in design is the duplication (or more) of particular components of a system for the purpose of increasing the overall system reliability. Redundancy in effect seeks to: (1) limit the impact of a single component with low reliability on the overall system reliability; (2) improve the reliability of a critical component in the system; and it does so by creating a virtual equivalent component of greater reliability than the single component in question." (Hoepfer et al., 2009).

The initial formulation (1978) referred to “intrinsically” safe plants, in contrast to “extrinsically” safe plants in which hazards were controlled by “extrinsic” protective equipment and safety features, instead of “intrinsically” safer processes. This distinction however was not maintained in other works by the author and others, and Kletz later replaced “intrinsically” with the now more common expression “inherently” safe.

Note that the inherent safety design principle is also discussed in Möller and Hansson (2008) as a broad category under which several safety principles are subsumed. The authors explain that this principle entails “that potential hazards are excluded, not just enclosed or otherwise coped with. [For example] fireproof material are used instead of inflammable ones […] and this is superior to using inflammable material but keeping temperatures low.”

Kletz further developed the inherent safety design principle and identified several pillars for ways of achieving it. These pillars include what the author refers to as intensification, substitution, and attenuation. Table 1 provides a brief description of these and other pillars of the inherent safety principle. The reader interested in more details is referred to Kletz and Amyotte (2010), Goraya *et* al., (2004), Khan and Amyotte (2003), and Bollinger and Crowl (1997).

Table 1. Pillars of the inherent safety design principle. Adapted form Kletz and Amyotte (2010) and Khan and Amyotte (2003)

Principle	**Description**
Process Intensification (PI) and minimization	Use smaller quantities of hazardous material and/or perform a hazardous procedure as few times as possible
Substitution	Replace hazardous materials/processes with safer ones
Attenuation	Use hazardous materials in their least hazardous forms and/or operate the system at comparably safer operating conditions
Limitation of Effects	Opt for changes in the process design with less severe effects
Simplification	Avoid complexities in the process design and eliminate excessive use of add-on safety features and protective devices

A careful examination of the entries in Table 1 shows similarities between the pillars of the inherent safety principle and Haddon’s safety strategies. For example, the reduction of the amount of energy contained in the process closely resembles *Intensification*, and the reduction of the rate of release of energy is similar to *Attenuation*.

The difference in framing various (overlapping) safety principles by different authors is unsurprising, and it reflects to some extent their particular background and interests. Kletz's work, for example, as noted previously, is grounded in the chemical industry, and is best understood and readily applicable in that industry. It remains nonetheless relevant for other hazardous industries. The inherent safety design principle and its pillars will be subsumed in part under the *un-graduated response* principle in Section 3.

2.4. Managerial and organizational safety principles or guidelines

Although beyond the scope of the present work (with its engineering focus on safety principles), it is worth noting that an important literature exists and addresses *organizational* safety principles or guidelines. The literature covers what is knows as High Reliability Organizations (HRO), and it empirically examines what successful organizations do—how they organize and manage hazardous systems and processes—to promote and ensure system safety. The reader interested in this line of inquiry is referred to the excellent work by Weick and Sutcliffe (2007) for a synthesis of the HRO literature.

3. System safety principles

Our proposed set of safety principles follows in the spirit of the works discussed in Section 2, and in some cases it builds and expands on them. The principles are first presented at some level of abstraction, which leaves them domain-independent and broadly applicable across industries. Then some of their practical aspects are highlighted and examples are provided to illustrate their implementation in specific contexts.

We relate our safety principles to the notions of hazard level, accident sequence, and conditional probabilities (of further hazard escalation or advancement of an accident sequence). Note that while we provide some simple formal representations of these safety principles and their consequences using probability notation and the Discrete Event Systems (DES) formalism[5], these representations are not necessary for the comprehension of the principles. The reader not familiar with such formalisms may skip the equations, and this will not compromise in any way his or her understanding of the safety principles.

[5] Formal representation can provide additional precision in defining the safety principles and their consequences (beyond a textual description).

The notions of accident sequence and hazard level, which are briefly reviewed next, can help further illuminate the purpose and consequences of these principles as will be seen shortly.

3.1 Background information: accident sequence and hazard level

An accident sequence can be represented in the form of an event tree, starting with an off-nominal initiating event and terminating in the accident state—the uncontrolled release of energy and its consequences. For example, Figure 1 shows a simplified version of an Event Tree Analysis for a generic nuclear reactor. The initiating event here considered is the break of the main coolant pipe.

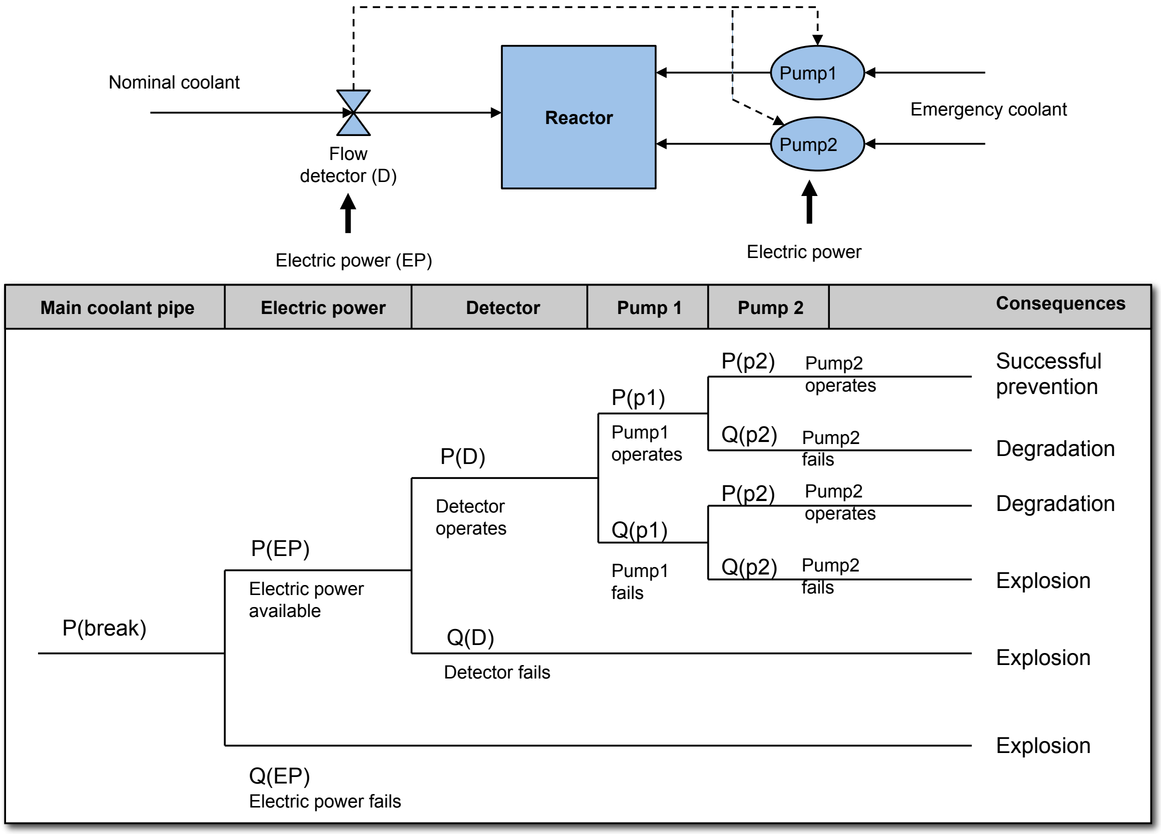


Figure 1. Simplified Event Tree Analysis for a reactor following the break of main coolant pipe (the initiating event). P = probability of success; Q = 1 – P = probability of failure. Adapted from Billington and Allan (1992).

The event tree reads from left to right. For example, in the path leading to the fourth consequence from the top (explosion), we have the following events: the main coolant pipe breaks; electric power is available upon demand to support the activation of the flow detector and emergency pumps; the flow detector operates properly and detects loss of main coolant; information is conveyed to activate redundant emergency pumps; pump 1 fails to activate; pump 2 also fails to activate, and this sequence of events leads to the explosion. The event tree can be further expanded to examine more possibilities and add further resolution to the consequences of the explosion and other branches (Saleh et al., 2013).

For our purposes, we will note more generally that an accident sequence can be represented by the concatenation of a series of events (denoted by the letter "e"), starting from an off nominal initiating event (denoted by "IE") and leading to an accident (denoted by "A"), as shown in Eq. 1 and Figure 2. Each event "e" presents two subscripts: the first one identifies its position inside the string **s**, while the second one identifies the initiating event. Event $\mathbf{e}_{2,1}$ defines an event that appears as a second link in a string **s** and that follows the initiating event IE_1. Notice that more accidents correspond to each initiating event, and that different initiating events can lead to the same accident unfolding. For simplicity, in Figure 2 we numerated the accidents starting from the top one as A_1. The string **s** also has two subscripts; the first corresponds to the initiating event, and the second to the final accident state. For example, Eq. 1 shows the accidents sequence represented by the string $s_{1,k}$, which starts with the initiating event IE_1 and terminated in the accident state A_k:

$$\boldsymbol{s}_{1,k} = IE_1 \boldsymbol{e}_{2,1} \boldsymbol{e}_{3,1} \ldots \boldsymbol{e}_{n,1} A_k \qquad (1)$$

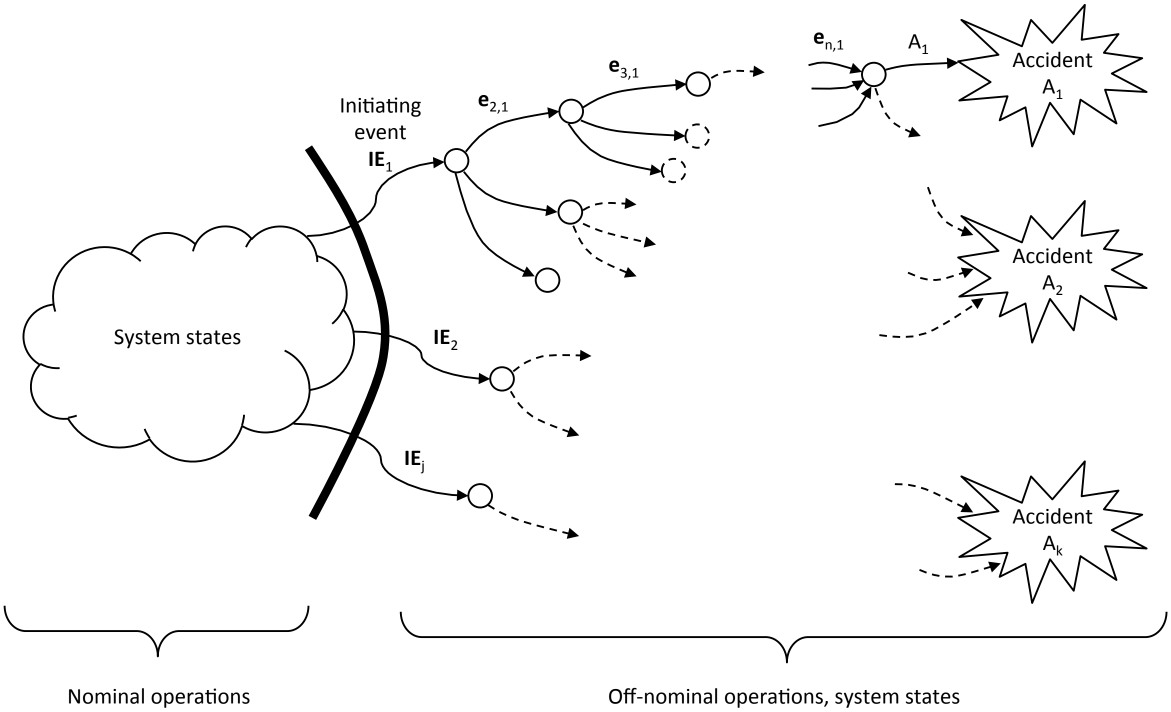


Figure 2. Illustrative example of the concept of accident sequence, with propagation of initiating events to accident states (Saleh *et* al., 2013)

For simplicity, we will occasionally drop the second subscript of an event **e**, and only index it with respect to its position in a given string as $\mathbf{e}_i$ (the ith event in an accident sequence).

Equation 1 is based on the mathematical framework of Discrete Event System (DES). The specifics are not relevant for our purposes (for details, see for example Cassandras and Lafortune, 2008). The important point is the way in which an accident sequence can be represented, namely as a string (denoted by the letter "***s***") of events and with multiple possible paths between different initiating events and accident states.

The conditional probability of accident A_k occurring given the occurrence of the initiating event IE_i can be written as follows:

$$p\left(A_k \mid IE_i\right) \tag{2}$$

This conditional probability is the sum over all paths starting from IE_i and leading to A_k. At a *local* level, given that an accident sequence has been initiated, the conditional probability that it will further advance or escalate is expressed as follows:

$$p\left(e_{i+1} \mid e_i\right) \tag{3a}$$

Or more generally:

$$p\left(e_k \mid e_i\right) \quad \text{for } k > i \tag{3b}$$

The idea of an accident sequence and the conditional probabilities associated with its escalation can help define or intuitively convey the notion of hazard level (*H*). Intuitively, the hazard level can be conceived of as the closeness of an accident to being released (Saleh *et* al. 2014). It is thus related to the extent an accident sequence has advanced: the further the sequence has escalated, the more hazardous the situation is. For example, using Eq. 1 and Figure 2, we can note:

$$H\left(IE_1 e_{2,1} e_{3,1} e_{4,1}\right) > H\left(IE_1 e_{2,1}\right) \tag{4}$$

For the situation in the left-hand side of Eq. 4, more adverse conditions are aligned and more events in the accident sequence have occurred than the situation in the right-hand side. The left-hand hazard level in the system or plant is thus higher and the accident is closer to being released. Figure 3 shows a typical example of a relation between an accident sequence and the dynamics of hazard escalation. In this case, only one string and one outcome are shown (a generic accident A).

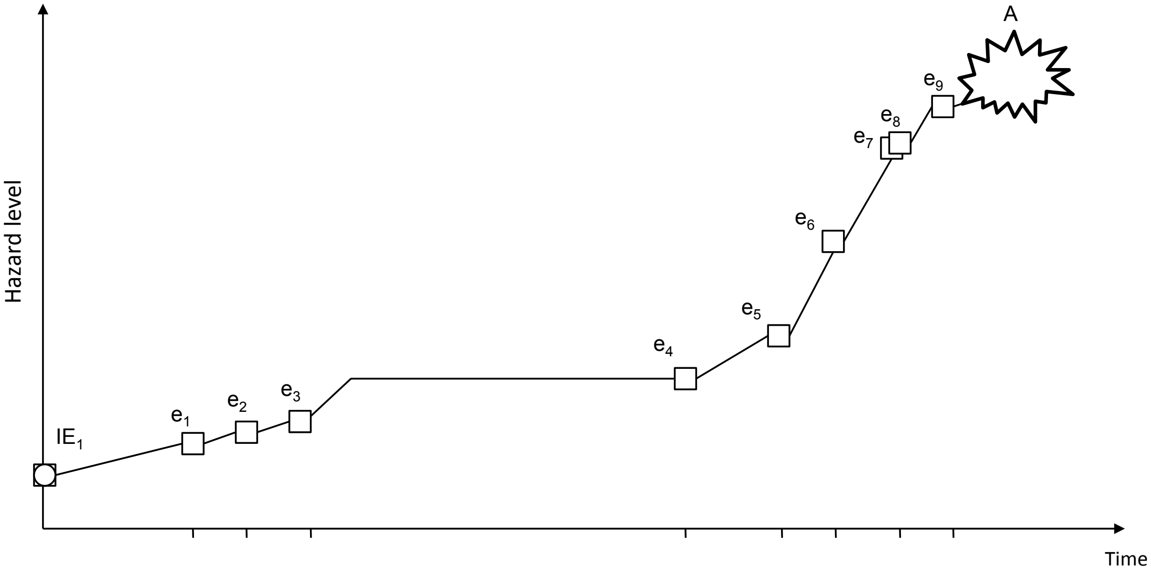


Figure 3. **Illustrative example of an accident sequence and hazard level escalation over time**

The operation of a hazardous process or system involves the management and handling of the dynamics of its hazard level. The dynamics of hazard escalation can be both time driven and event driven, and all else being equal, the hazard level scales with the extent of potential adverse consequences (PAC). We indicate this functional dependency as follows:

$$H = H(t,\ \mathrm{e},\ PAC) \tag{5}$$

The conditional probabilities previously mentioned can also be added to the expression in Eq. 5. They are in its current form implicit in the string of events (*e*) of an accident sequence. Note that the potential adverse consequences are a function of both the amount of energy involved or being handled, and the extent of vulnerable resources in its neighborhood (people and structures). For example, a chemical plant in the middle of a densely populated city has a higher potential for adverse consequences than if it were sited in a remote industrial zone.

These concepts, accident sequence, conditional probabilities of sequence escalation, and hazard level, will be referred to next when discussing the safety principles. They will help us illustrate for example the effects of these principles on the advancement of an accident sequence and on the dynamics of hazard escalation, as we will see shortly.

3.2 The fail-safe safety principle

Consider a function performed or implemented by a particular item in a system. The failure of this item or disruption/termination of its function can propagate and affect the system in different ways. For example it can lead to a cascading failure (domino effect), which would result in a complete system failure or accident (e.g., nodes in an electric power grids operating at maximum capacity). It can also remain confined to the neighborhood of the failed item and have a limited impact at the system level.

The fail-safe principle imposes, or is defined by, one particular solution to the problem of how a local failure affects the system level hazard. Specifically, the fail–safe principle requires that the failure of an item in a system or disruption/termination of its function should result in operational conditions that (i) block an accident sequence from further advancing, and/or (ii) freeze the dynamics of hazard escalation in the system, thus preventing potential harm or damage.

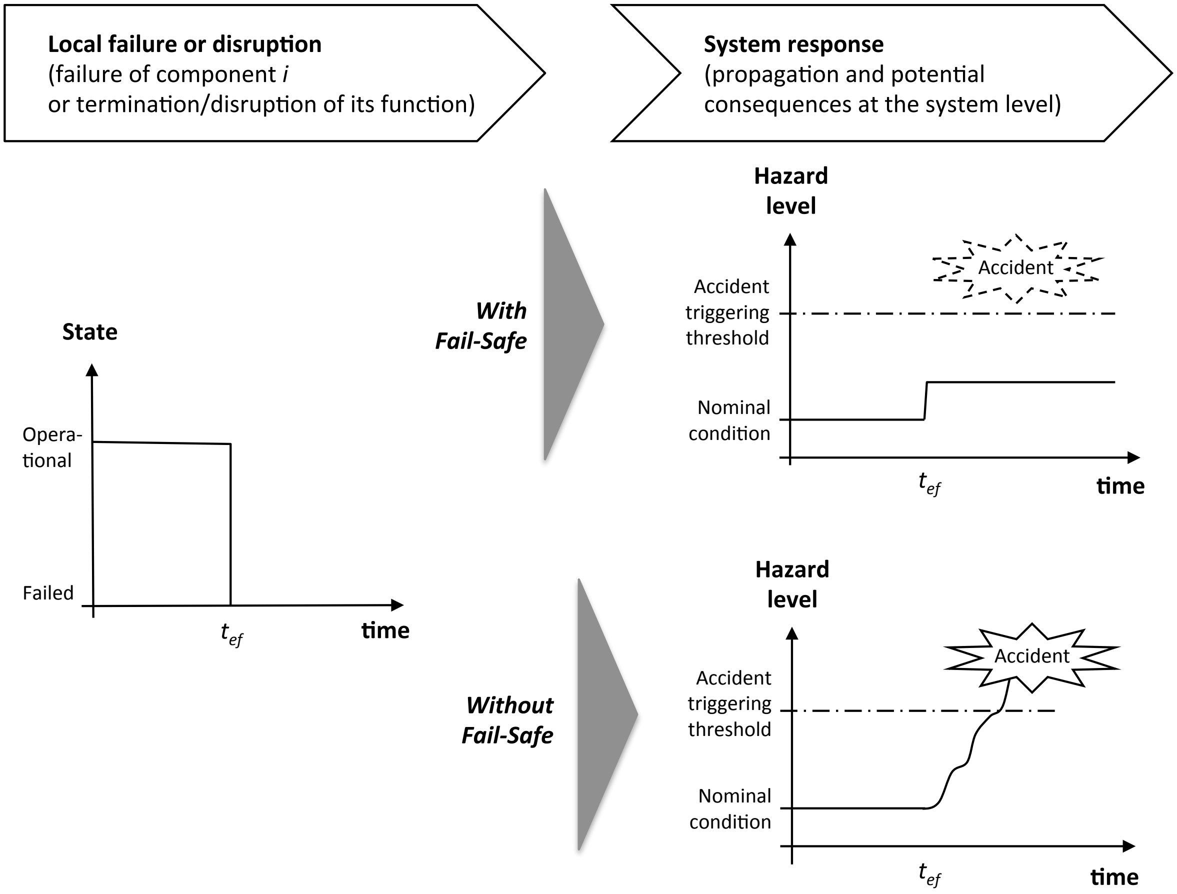
In light of the concepts introduced in subsection 3.1, the effects of the fail-safe principle can be expressed as follows:

e_f : failure of the item/function of interest at time t_{e_f}

$$\begin{cases} \dfrac{\partial H}{\partial t} = 0 \quad \text{for } t > t_{e_f} \\ and \\ p\left(e_{f+k} \mid e_f\right) = 0 \qquad e_{f+k} \in \boldsymbol{s} \ \text{ following } e_f \end{cases} \tag{6}$$

Eq. 6 expresses the fact that the dynamics of hazard escalation are frozen after the failure of the item/function, and the accident sequence is blocked (see Fig. 4).

Conversely, if the fail-safe principle is not implemented, the item's failure, or termination of the function it performs, would aggravate a situation by further escalating its level of hazard, thus initiating an accident sequence or leading to an accident, as shown in Figure 4. For example, air brakes on trains and trucks are maintained in the open position by pressure in the lines; should the pressure drop because of leakage or any other failure mechanism, the brakes will be applied. A similar mechanism exists in elevators: a spring force activated electrically holds the brakes in the open position. In the event of a power failure, the brakes automatically engage. The difference between the brakes failing in the open position and leading to the free-fall of the elevator, and the brakes failing in the engaged position thus preventing a hazardous situation from unfolding, is the result of a creative implementation of the fail-safe principle in this particular situation. Popular legend notwithstanding, the only accidents involving elevators falling have occurred when the building itself has been catastrophically damaged (Paumgarten, 2008).



Local failure or disruption
(failure of component *i*
or termination/disruption of its function)
System response
(propagation and potential
consequences at the system level)
State
Opera-
tional
Failed
t_{ef}
time
With Fail-Safe
Without Fail-Safe
Hazard level
Accident
Accident triggering threshold
Nominal condition
t_{ef}
time
Hazard level
Accident
Accident triggering threshold
Nominal condition
t_{ef}
time

Figure 4. Illustrative comparison of system behavior over time following a local failure, both with the implementation of the fail-safe principle and without it (t_{ef} is the time of occurrence of the failure of the component/function of interest)

Another example of the implementation of the fail-safe principle is the "dead man's switch" for train operators: should they fall asleep or become unconscious, the device is no longer held down, and as a result the brakes are applied. A similar device is used in chainsaws, snowmobiles, jet skis, and during aircraft refueling (the activity is stopped). More complex implementations of the fail-safe principle can be found in nuclear reactors where self-shutdown is initiated if critical operating conditions are reached. While the details are not relevant for our discussion, the important idea is that the fail-safe principle can be implemented in a variety of ways and it requires engineering creativity and technical ingenuity to conceive and implement in various contexts (e.g., the US patent and trademark office lists more than 20,000 patents proposing various fail-safe mechanisms[6]). One final comment about this principle: while there may be situations or items for which the fail-safe principle is incompatible with their design or is simply not implementable, it is nevertheless important that this principle always be considered and carefully assessed in any design endeavor before it is ruled out.

3.3 The safety margins (or safety reserves) principle

The adoption of safety margins is a common practice in civil engineering where structures are designed with a safety factor to account for larger loads than what they are expected to sustain, or weaker structural strength than usual due to various uncertainties. The importance of safety margins for structures such as bridges and levees, which have to cope with the uncertainty of operational and environmental conditions such as wind force and wave height, is easy to understand. We propose that the idea of safety margins in civil engineering is an instantiation of a broader safety principle, which we will refer to by the same name. The safety margin principle extends beyond civil engineering and is more diverse in its implementation than the particular form it takes for structures. In other words, safety margins can take multiple forms and be adopted in a variety of contexts, as we will see shortly.

The safety margin principle has a simple form and is intuitively understood. It requires first an estimation of a critical hazard threshold for accident occurrence, $\hat{H}_{critical}$, and an understanding of the dynamics of hazard escalation in a particular situation. For example, methane in coalmines enters an "explosive range" when its concentration in the mine atmosphere reaches between 5% and 15% (Saleh and Cummings, 2011). Reaching the 5% threshold for example can be considered a critical hazard threshold in the mine.

The safety margin principle requires that features be put in place to maintain the operational conditions and the associated hazard level at some "distance" away from the estimated critical hazard threshold or accident-triggering threshold. This distance or norm is the general form of a safety margin (*sm*), and it can be expressed as follows:

$$\left\| \hat{H}_{critical} - H_{op}(t) \right\| \geq sm \tag{7a}$$

or in relative terms:

[6] U.S. Patent and Trademark office, full-text and image database after 1976, at http://patft.uspto.gov/netahtml/PTO/search-bool.html [accessed May 7, 2013]

$$\left\| \frac{\hat{H}_{critical} - H_{op}(t)}{\hat{H}_{critical}} \right\| \geq sm_{\%}$$
$$or \tag{7b}$$
$$\left\| H_{op}(t) \right\| \leq \frac{\left\| \hat{H}_{critical} \right\|}{1 + sm_{\%}}$$

For instance, in the coal mine example, a safety margin can be established with respect to the risk of methane explosion by maintaining methane concentration below say 3% in the mine atmosphere, 2 percentage points below the critical hazard level. The difference between the operational upper limit and the boundary of the explosive range, the triggering threshold, is a particular form of safety margin in this context. This safety margin can be established through reliable methane monitoring throughout the mine, and most importantly through proper ventilation. Unfortunately, most mine explosions worldwide are the direct result of a violation of this basic safety principle or its flawed implementation (Saleh and Cummings, 2011).

In a different context, pressure safety valves are another form of implementation of the safety margin principle in the oil and gas and chemical industries. If P_{crit} is the critical pressure beyond which the structural integrity of a containment vessel is compromised, the safety relief valve is triggered when the inequality below is reached:

$$P(t) \leq \frac{P_{crit}}{1 + sm_{\%}} \tag{8}$$

The purpose of the valve is to maintain the pressure in the vessel at some *distance* away from the accident-triggering pressure threshold. The accident in this case would be loss of containment, which might result in fire, explosion, and/or release of toxic material into the environment.

Figure 5 provides a visual representation of the safety margins principle.

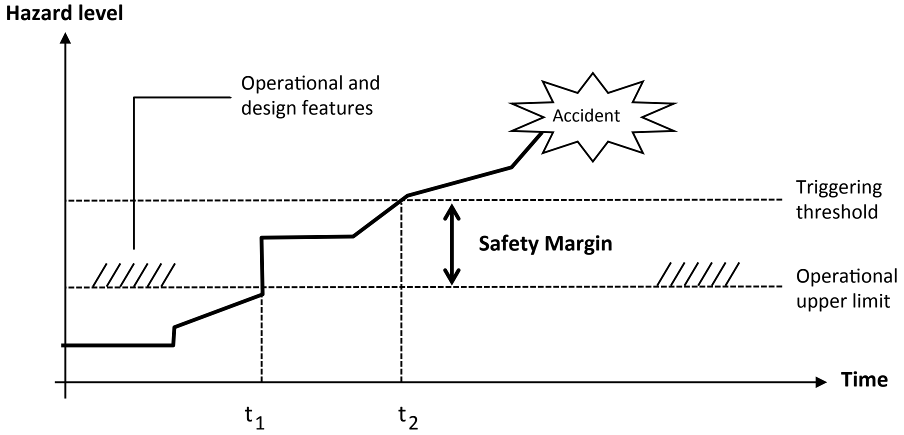


Figure 5. Illustration of the safety margins principle with a sample accident trajectory from a nominal operating condition to an accident. A larger margin makes it more likely that the system state will not reach the accident-triggering threshold, or that a longer time window is available to detect a system state that has crossed the operational upper limit (for nominal conditions) and abate the hazardous situation before an accident is triggered.

The justification for this principle is best understood in light of the argument put forth by Möller and Hansson (2008) as discussed in subsection 2.2, namely that one objective of all safety principles is the reduction in uncertainty about the occurrence of an adverse event (both epistemic and aleatory uncertainties). In the case of the safety margins principle, notice that $\hat{H}_{critical}$ is the **estimated** critical hazard threshold or accident-triggering threshold. The actual threshold is unknown or best modeled as a random variable. Safety margins are one way for coping with uncertainties in both the critical hazard threshold and in our ability to manage the operational conditions in a system such that their associated hazard level $H_{op}(t)$ does not intersect with the real but unknown $H_{critical}$.

Finally, we note that as with the fail-safe principle, the safety margins principle can also be implemented in a variety of ways and it requires creativity and technical ingenuity to conceive and design in different contexts and for handling different types of hazards. It is also important to carefully assess the trade-offs before corners are cut and safety margins are shrunk.

3.4 The un-graduated response principle: rules of engagements with hazards

The use of force in a military or law enforcement context is governed by a set of Rules of Engagements or Rules for the Use of Force (CJCSI, 2005). The principal tenet of these rules is that of a *graduated response*, namely that if force is deemed necessary, it ought to be applied gradually in relation to the extent of a demonstrated belligerence, as a last resort, and only the minimum force necessary to accomplish the mission should be used.

The opposite of this tenet holds for dealing with safety issues, and the corresponding principle we refer to as the un-graduated response or rules of engagements with technological hazards. This principle for accident prevention and mitigation articulates a hierarchy of preferences for safety interventions.

The un-graduated response principle is closely related to and overlaps with other safety principles, especially defense-in-depth discussed next. But it is worth examining separately as it articulates an important safety idea in a novel and forceful way. Given that the objective of the present work is mainly educational (to be used in safety training and education), examining this idea from different perspectives for emphasis is worthwhile.

Kill first: The un-graduated response principle posits that the first course of action to explore for accident prevention and mitigation is the possibility of eliminating a hazard all together. We refer to this attitude as "kill first" or use creativity and technical ingenuity as a first resort to eliminate the hazard, regardless of the extent of its *belligerence* (lethal use of force against hazards).

For example, many precautions can be taken when transporting hazardous materials, such as the use of thicker and sturdier containers. But eliminating the hazard all together instead of better containing it, by transporting a safer substitute for example ought to be the first course of action to consider and examine for feasibility. Similarly if a heat source or electric wires are in the vicinity of flammable material, the hazard can be controlled or the probability of an accident reduced by using proper wire isolation and placing the wires within fireproof protective jackets. But this particular hazard, the co-location of the electric wires and flammable material, can be eliminated by re-routing the wires through another location—the preferred course of action by virtue of this safety principle.

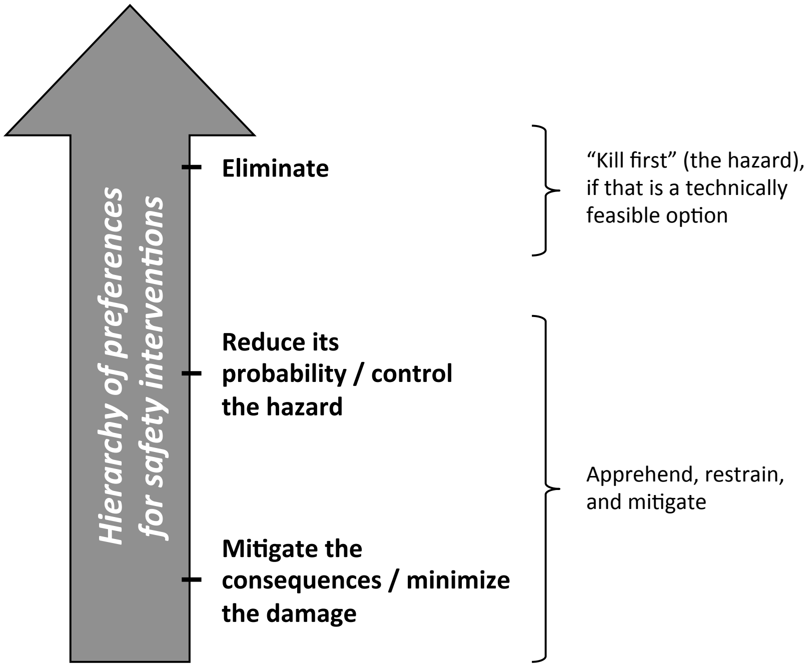

Hierarchy of preferences
for safety interventions
Eliminate
Reduce its probability / control the hazard
Mitigate the consequences / minimize the damage
“Kill first” (the hazard), if that is a technically feasible option
Apprehend, restrain, and mitigate

Figure 6. Illustration of the hierarchy of preferences for safety interventions by virtue of the un-graduated response principle (not to be considered with an *exclusive or*)

Next, if the hazard cannot be eliminated, the second course of action is to control it or reduce its likelihood of escalating into an accident. Figuratively, if “kill first” is not feasible, then proceed to “apprehend and restrain”. A third and concurrent course of action is to devise ways to mitigate the consequences or minimize the damage should the hazard escalate into an accident (Figure 6). We revisit these issues in more details when we discuss the defense-in-depth safety principle next. Notice that the un-graduated response principle covers the general idea in Kletz’ inherent safety principle discussed previously, and the “safety by design” concept (Bollinger, 1996; Khan and Amyotte, 2003; Kletz and Amyotte, 2010).

3.5 The defense-in-depth principle

Defense-in-depth is a fundamental safety principle and one whose importance cannot be underestimated. We believe this principle should be central to the education of all engineers and anyone interested in system safety issues.

This principle targets all the elements of an accident sequence (Eq. 1), and it *intervenes* with all the arguments of the hazard level function (Eq. 5), as we will see shortly.

Defense-in-depth derives from a long tradition in warfare by virtue of which important positions were protected by multiple lines of defenses (e.g., moat, outer wall, inner wall). First conceptualized in the nuclear industry, defense-in-depth became the basis for risk-informed decisions by the U.S. Nuclear Regulatory Commission (NRC, 2000; Sørensen et al., 1999-2000), and it is adopted under various names in other industries. Defense-in-depth has several pillars:

i. Multiple lines of defenses or safety barriers should be placed along potential accident sequences;

ii. Safety should not rely on a single defensive element (hence the "depth" qualifier in defense-in-depth);

iii. The successive barriers should be diverse in nature and include technical, operational, and organizational safety barriers. In other words, defense-in-depth should not be conceived of as implemented only through physical defenses.

The various safety barriers have different objectives and perform different functions. The first set of barriers, or line of defense, is meant to prevent an accident sequence from initiating. In light of the previous discussion of an accident sequence (subsection 3.1), the first line of defense implies that safety features are devised and put in place such that the probability of an accident-initiating event (IE) is minimized:

$$\min\left[p\left(IE_i\right)\right] \tag{9}$$

Should this first line of defense fail in its **prevention** function, a second set of safety defenses should be in place to block the accident sequence from further escalating:

$$\min\left[p\left(e_{i+k} \mid e_i\right)\right] \qquad \forall i,k \quad for \quad e_i \in \boldsymbol{s} \quad \text{and} \quad e_{i+k} \in \boldsymbol{s} \quad \text{following } e_i \tag{10}$$

Finally should the first and second lines of defense fail, a third set of safety defenses should be in place to **contain the accident and mitigate its consequences**. This third line of defense is designed and put in place based on the assumption that the accident will occur, but its potential adverse consequences (PAC) should be minimized. Lifeboats are one illustration of this third line of defense. The objective of the third line of defense can be expressed as follows:

$$\min\left(PAC \mid A\right) \tag{11}$$

These three lines of defenses constitute defense-in-depth and its three functions, namely (i) prevention, (ii) blocking further hazardous escalation, and (ii) containing the damage or mitigating the potential consequences. Figure 7 illustrates this safety principle, along with a particular accident sequence.

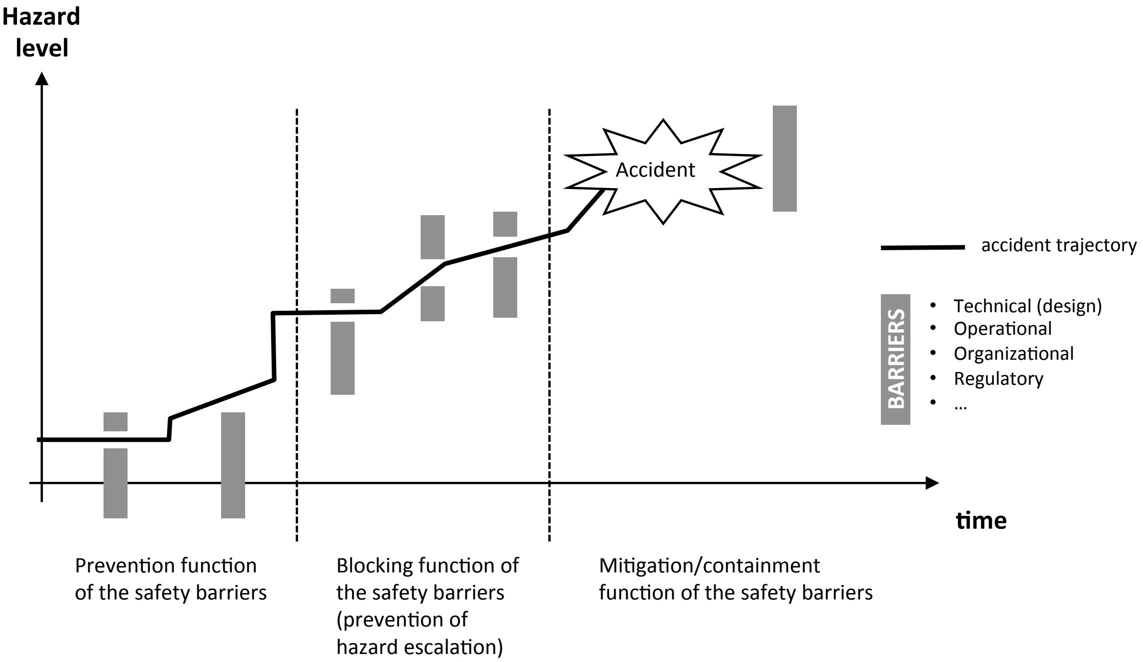

Hazard level
Accident
accident trajectory
BARRIERS
• Technical (design)
• Operational
• Organizational
• Regulatory
• …
time
Prevention function of the safety barriers
Blocking function of the safety barriers (prevention of hazard escalation)
Mitigation/containment function of the safety barriers

Figure 7. Illustration of the defense-in-depth safety principle, along with a hypothetical accident sequence (its occurrence is the result of the absence, inadequacy, or breach of various safety barriers)

Accidents typically result from the absence, inadequacy, or breach of defenses, or the violation of safety constraints, as illustrated in Figure 7 (Rasmussen, 1997; Svedung and Rasmussen, 2002; Leveson, 2004). It is interesting to note that the U.S. Department of Energy defines an accident as an "unwanted transfer [or release] of energy that, due to the absence or failure of barriers and controls, produces injury to persons [or] damage to property" (DOE, 1997). This view is related Haddon's energy basis of accidents and injuries, and it provides a useful bridge between defense-in-depth and Haddon's safety strategies discussed in subsection 2.2. The understanding of defense-in-depth can thus be expanded and this principle be viewed as functioning to prevent, contain, and limit unwanted releases of energy.

The notion of a safety barrier is the embodiment of the "defense" part of defense-in-depth in the sense that defenses are realized through barriers deliberately inserted along potential accident sequences and prior to their initiating events.

It can be seen that the previous safety principles overlap to some extent with defense-in-depth. For example, the implementation of a fail-safe mechanism, or the establishment of a safety margin, can be considered as different forms of barriers in the layout of defense-in-depth. And the un-graduated response principle reflects to some extent the different functions of the multiple lines of defenses (their "depthness"). This overlap is useful, and it provides us with an opportunity to emphasize certain foundational ideas in safety education.

Finally, as with the previous safety principles, defense-in-depth can be implemented in many ways and it requires significant ingenuity—technical, operational, organizational, and regulatory—to conceive and implement in a variety of contexts and for dealing with different types of hazards and uncertainties.

3.6 The observability-in-depth principle

Observability-in-depth plays a distinctive role in system safety, and it contributes to accident prevention in a fundamentally different way than the previous principles, as we will discuss shortly. Observability-in-depth brings an online real-time mindset to accident prevention (i.e., during system operation)—an aspect that was either missing or not explicitly recognized in the previous principles. Unlike defense-in-depth and its relation to the energy model of accidents, observability-in-depth is fundamentally an information-centric principle, and its importance in accident prevention is in the value of the information it provides and actions or safety interventions it spurs, as we will see shortly.

Observability-in-depth does not affect or intervene directly in an accident sequence, unlike the previous principles, but it scans for accident pathogens and monitors for hazard escalation and advancement of accident sequences. Its importance cannot be underestimated, and its significance is best motivated by first considering situations in which this principle was NOT implemented. Violations of the observability-in-depth principle highlight not the causal chain of an accident sequence—why the accident happened—but the causal factors that failed to support accident prevention—why blocking the accident sequence did not happen.

There are several mechanisms in the design of complex systems that can contribute to concealing the occurrence of hazardous events (e.g., redundant component failures or build-up of accident pathogens/latent failures) and the transition of the system to an increasingly hazardous state, which make "systems more […] opaque to the people who manage and operate them" (Reason, 1997). As a result, system operators may be left blind to the possibility that hazard escalation is occurring, thus decreasing their situational awareness and shortening the time they have to intervene before an accident is released. In other words, these safety blind spots translate into a shrinking of the time window available for operators to identify an unfolding hazardous situation and intervene to abate it. Several accident reports identified hidden failures and unobservable accidents pathogens as important contributing factors to the accidents, the Three Mile Island and the Texas City refinery accidents being representative cases (Hopkins, 2001; Saleh et al., 2013).

To visually illustrate this argument, consider the situation represented in Figure 8. This is similar to the dynamics of hazard level and accident sequence represented previously in Figure 3, except we now distinguish between the actual hazard level, *H(t)*, and the estimated or assumed hazard level, $\hat{H}(t)$.

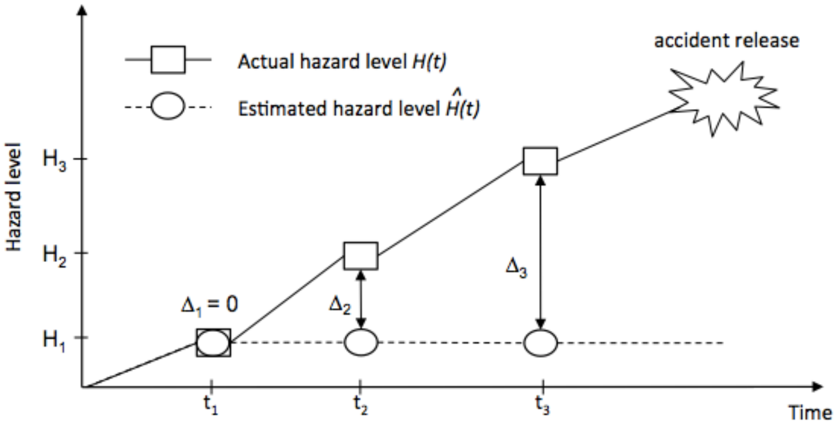


Figure 8. Hazard escalation over time and the violation of the observability-in-depth principle (Saleh et al., 2014). The figure shows how underestimating the actual hazard level (ovals) can lead to an accident occurring seemingly without warning (rectangles). The gap between the these two quantities (Δ) represents a loss of situational awareness and can result in the operators making flawed decisions or not taking a needed safety action, which in turn can compromise the safe operation of the system or fail to check the escalation of an accident sequence.

Roughly speaking, operators make decisions during system operation, which are based on and affect the hazard level in a system. If the system conditions/states are not carefully monitored and reliably reported, there is a distinct possibility that the hazard level *estimated* by the operators will diverge from the *actual* hazard level reached by the system:

$$\Delta H \equiv \left\| \hat{H} - H\left(t\right) \right\| \tag{12}$$

The gap between these two quantities (Eq. 12) can result in the operators making flawed decisions, which in turn can compromise the safe operation of the system or fail to check the escalation of an accident sequence (e.g., no action when an intervention is warranted). The Three Mile Island and the Texas City refinery accidents as noted previously are examples of such situations. Details of these accidents and other similar cases can be found in (Saleh et al., 2014; Favarò and Saleh, 2013).

The discussion so far has considered, as a way of motivating the observability-in-depth principle, the safety implications when this principle is not implemented. We are now ready to examine this principle and its objectives.

Observability-in-depth is characterized by the set of provisions, technical (by design) and operational, designed to enable the monitoring and identification of emerging hazardous conditions and accident pathogens. Observability-in-depth requires that all safety-degrading events or states that safety barriers are meant to protect against be observable. It implies that various features be put in place to observe and monitor for the system state and breaches of any safety barrier, and reliably provide this feedback to the operators (Bakolas and Saleh, 2011; Favarò and Saleh, 2013). In light of Figure 8 and Eq. (12), it can be said that observability-in-depth seeks (i) to minimize the gap between the actual and the estimated hazard levels, and (ii) to ensure that at the hazard levels associated with the breaching of various safety barriers, H_1, H_2, and H_3 in the figure, the two curves (actual and estimated) coincide[7]. This concept can be expressed as follows:

$$\mathrm{e}_{b_i} : \text{ breach of safety barrier } b_i$$

$$\begin{cases} \min \Delta H \;\; \Leftrightarrow \;\; \min \left\| \hat{H} - H(t) \right\| \\ \mathit{and} \\ \Delta H_{b_i} = 0 \qquad \forall i \end{cases} \tag{13}$$

The **depth** qualifier in observability-in-depth has both a causal and a temporal dimension, and it characterizes the ability to identify adverse states and conditions far upstream (early) in an accident sequence (see Eq. 1). It also reflects the ability to observe emerging accident pathogens and latent failures before their effect becomes manifest on the system's output or behavior, or before an increasingly hazardous transition occurs in an accident sequence. To illustrate this point, consider Figure 9, which represents a set of safety barriers and

[7] Observability-in-depth echoes the old Russian say, "trust but verify". If the previous safety principles are meant to build (some) trust in the safe operation of a system, observability-in-depth is concerned with the "verify" part.

various hazardous states. Each hazardous transition/escalation in an accident sequence has a set of underlying causes; Figure 9 includes only one such set for readability purposes, the underlying causes of a transition from S_i to S_j in the form of a Fault Tree.

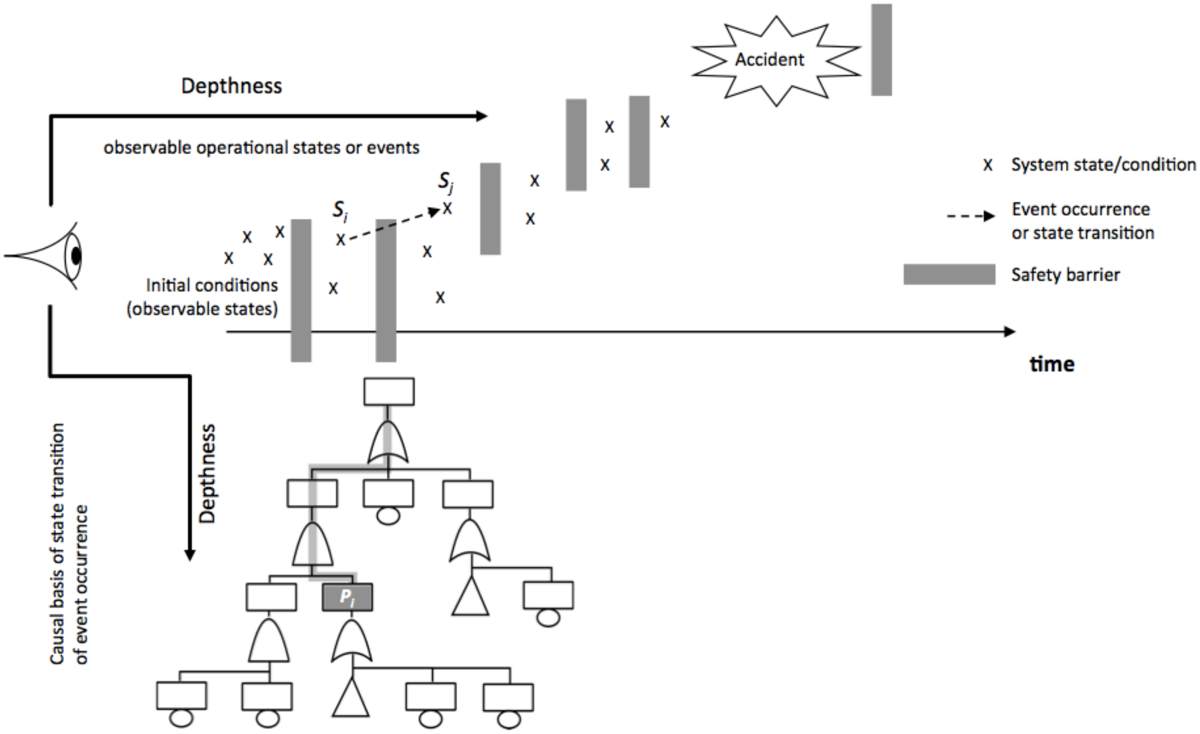


Figure 9. Illustration of the Observability-in-Depth principle. The figure shows (i) above the time axis the manifestation of the accident sequence and escalation of the hazard level as safety barriers are breached, and (ii) below the time axis the causal basis (why) of hazardous state transition. The "depthness" of observability (the two arrows in the figure) characterizes both the ability to *see* breaches of safety barriers and to identify accident pathogens (deeply buried) in the causal basis of a hazardous transition (before their effect is manifest on the system's output or behaviour).

The condition P_i in the fault tree is a latent failure or accident pathogen (Saleh et al., 2013); it does not have a visible effect on the system behaviour or operation until the second condition in its AND gate occurs. If the system reaches state S_i, the hazardous transition to S_j will occur, thus further advancing the accident sequence. The ability to observe such causal factors or accident pathogens in an accident sequence before they have a visible effect on the system operation is one measure of the *depthness* of observability. The other measure is that no line of defense should conceal the fact that the system has breached any one safety barrier and has reached a hazardous state the engineers and system designers meant to protect against.

Probability Risk Assessment (PRA) has traditionally been performed offline and used as a static tool to help identify and prioritize various risks before the system is in operation. Observability-in-depth introduces an online (real-time) mindset into risk analysis and management, and it supports the development of a "living" or online quantitative risk assessment. Further, it is worth clarifying that observability-in-depth is an important complement to defense-in-depth: the former prevents the latter from devolve into a defense-blind safety strategy, and the latter along with PRA, guide the establishment of provisions for monitoring safety functions and barriers. In short, observability-in-depth can help conceive of a **dynamic defense-in-depth safety strategy** in which some defensive resources, safety barriers and others, are prioritized and allocated dynamically in response to emerging risks (Bakolas and Saleh, 2011; Favarò and Saleh, 2013).

Finally, we note that, as with all the previous safety principles, observability-in-depth can be implemented in many ways, and it requires technical ingenuity to design and implement in a variety of contexts and for monitoring different types of hazards and states of safety barriers. As a final remark, the general approaches to fault detection and diagnosis in dynamical systems (Venkatasubramanian *et* al., 2003a, b, c) are subsumed under this principle and they constitute in some case specific forms of implementation of this principle (tactics).

4. Conclusion

As noted in the Introduction, risk analysis has been described as addressing three main questions:

(1) What can go wrong?
(2) How likely it is?
(3) What would be the consequences?

The safety principles examined in this work provide guidelines and conceptual support during system design and operation for addressing the most important follow-up question, namely:

(4) What are you going to do about it [what can go wrong]? Or how are you going to defend against it?

The observability-in-depth principle was distinctive in that it addressed the real-time version of the first question of risk analysis:

(Real-time mindset) What is going wrong, if anything? And how will you know it is happening?

The high-level safety principles discussed in this work, fail-safe, safety margins, un-graduated response, defense- and observability-in-depth, are domain-independent, technologically agnostic, and broadly applicable across industries. While no claim to exhaustiveness is made, we believe many detailed safety measures (tactics) derive from or relate to these principles. The translation of these safety principles into specific design features and safety measures requires, as was emphasized throughout this work, detailed knowledge of the system under consideration as well as creativity and technical ingenuity to conceive and implement in various context and for handling different risks.

We also related the safety principles examined in this work to the notions of hazard level, accident sequence, and conditional probabilities of further hazard escalation or advancement of an accident sequence. These notions helped us better illuminate the purpose and consequences of these principles. The principles were also related to the energy model of accidents (Haddon's contribution) and the broadly known "inherent safety principle" in the chemical industry (Kletz's contribution).

The main purpose of this work, as noted in the Introduction, is educational and it is meant to serve a useful role in safety training and education. We believe these principles are a useful addition to the intellectual toolkit of engineers, decision-makers, and anyone interested in safety issues, and they can provide helpful guidelines during system design and risk management efforts. Subsequent courses or safety training can build on this basis and examine detailed measures in the specific industry of interest to different audiences.

References

Apostolakis, G. E. (2004). "How useful is quantitative risk assessment?". *Risk Analysis*, Vol. 24, Issue 3, pp. 515-520.

Bakolas, E., and Saleh, J. H. (2011). "Augmenting defense-in-depth with the concepts of observability and diagnosability from Control Theory and Discrete Event Systems". *Reliability Engineering & System Safety*, Volume *96*, Issue 1, pp. 184-193.

Bollinger, R. E., & Crowl, D. A. (1997). *Inherently safer chemical processes: a life cycle approach*. Wiley-AIChE.

Bollinger, R E., Clark, D. G., Dowell, A. M. , Ewbank, R. M., Hendershot, D. C. , Lutz, W. K. , Meszaros, S. I., Park, D. E. , and Wixom, E. D. (1996). "Inherently Safer Chemical Processes: A Life Cycle Approach" ed. D. A. Crowl. New York: American Institute of Chemical Engineers.

Cassandras CG. and Lafortune S. (2008) “Introduction to Discrete Event Systems”. Second edition, New York: Springer.

Chairman of the Joint Chiefs of Staff Instruction (CJCSI) 3121.01B (2005). “*Standing Rules of Engagement/Standing Rules for the Use of Force for U.S. Forces”.*

DOE. (1997). Implementation Guide for Use with DOE Order 225.1A, Accident Investigations, DOE G 225.1A-1. Washington, DC: US Department of Energy.

Favarò, F. M., and Saleh, J. H. (2013) “Observability in Depth: novel safety strategy to complement defense-in-depth for dynamic real-time allocation of defensive resources”. Presented at the ESREL Conference September 29 – October 2 2013, Amsterdam.

Gibson, J. J. (1964). “The contribution of experimental psychology to the formulation of the problem of safety brief for basic research” *In*: Haddon Jr, W, *et* al., (eds). Accident research: methods and approaches. Harper & Row, New York.

Goraya, A., Amyotte, P. R., & Khan, F. I. (2004). An inherent safety–based incident investigation methodology. *Process safety progress*, *23*(3), 197-205.

Haddon Jr, W. (1980a). “Advances in the epidemiology of injuries as a basis for public policy.” *Public Health Reports*, Volume 95, Issue 5, pp. 411–421.

Haddon Jr, W. (1980b). “The basic strategies for preventing damage from hazards of all kinds.” *Hazard Prevention*, Volume 16, Sept.–Oct., pp. 8–11.

Hopkins, A. (2001). “Was Three Mile Island a ‘Normal Accident’?”. *Journal of Contingencies and Crisis Management*, Volume *9,* Issue 2, pp. 65-72.

Hoepfer, V. M., Saleh, J. H., and Marais, K. B. (2009). “On the value of redundancy subject to common-cause failures: Toward the resolution of an on-going debate”. *Reliability Engineering & System Safety*, Volume 94, Issue 12, pp. 1904-1916.

ISO 31000. “Risk Management – Principles and Guidelines”. International Standard, ISO31000: 2009(E)

ISO 31010. “Risk Management–Risk Assessment Techniques.” International Standard IEC/FDIS 31010: 2009(E)

Kaplan, S. & Garrick, B. J. (1981). "On The Quantitative Definition of Risk". *Risk Analysis*, Volume 1, Issue 1, pp. 11-27.

Khan, F. I. and Amyotte, P. R. (2003). "How to Make Inherent Safety Practice a Reality". *The Canadian Journal of Chemical Engineering*, Volume 8, Issue 1, pp. 2–16.

Kletz, T. A. (1978). "What you don't have, can't leak". *Chemistry and Industry*, Volume 6, pp. 287-292.

Kletz, T. (1998). "Process plants: A handbook for inherently safer design*"*. Taylor & Francis.

Kletz, T., and Amyotte, P. R. (2010). "Process plants: a handbook for inherently safer design – 2nd edition". CRC Press.

Leveson, N. G. (2004). "A new accident model for engineering safer systems". *Safety Science*, Volume 42, Issue 4, pp. 237-270.

Möller, N., and Hansson, S. O. (2008). "Principles of engineering safety: risk and uncertainty reduction". *Reliability Engineering & System Safety*, Volume *93,* Issue 6, pp. 798-805.

NRC, US (2000). "Causes and Significance of Design Basis Issues at US Nuclear Power Plants". Draft Report, Washington, DC: US Nuclear Regulatory Commission, Office of Nuclear Regulatory Research.

Pate-Cornell, E., "Uncertainties in risk analysis: Six levels of treatment." Reliability Engineering and System Safety, Vol. 54, No. 2, 1996, pp. 95–111.

Paumgarten, N. (2008). "Up and then down – The lives of elevators". The New Yorker, April 28th 2008, http://www.newyorker.com/reporting/2008/04/21/080421fa_fact_paumgarten?printable =true&currentPage=all

Rasmussen, J. (1997). "Risk management in a dynamic society: a modeling problem". *Safety Science,* Volume 27, Issues 2-3, pp. 183–213.

Reason, J. T. (1997). "Managing the risks of organizational accidents". Aldershot, Hants, England; Brookfield, Vt., USA: Ashgate.

Rivara F. P., Cummings P., Koepsell T. D. Grossman D. C., Maier R. V. (2001) "Injury control". Cambridge University Press, Cambridge, UK.

Runyan, C. W. (2003) "Back to the future—revisiting Haddon's conceptualization of injury epidemiology and prevention." Epidemiologic Reviews, Volume 15, Issue 1, pp. 60–64.

Saleh, J. H., Marais, K. B., Bakolas, E., Cowlagi, R. V. "Highlights from the literature on accident causation and system safety: Review of major ideas, recent contributions, and challenges." Reliability Engineering and System Safety, Volume 95, Issue 11, 2010, pp. 1105–1116.

Saleh, J. H., and Cummings, A. M. (2011). "Safety in the mining industry and the unfinished legacy of mining accidents: Safety levers and defense-in-depth for addressing mining hazards". *Safety Science*, Volume 49*,* Issue 6, pp. 764-777.

Saleh, J. H., Haga, R. A., Favarò, F. M., Bakolas, E. (2014). "Texas City Refinery Accident: Case Study in Breakdown of Defense-In-Depth and Violation of the Safety-Diagnosability Principle". *Engineering Failure Analysis*, Volume 36, pp. 121-133.

Saleh, J. H., Saltmarsh, E., Favarò, F. M., Brevault, L. (2013). "Accident precursors, near misses, and warning signs: critical review and formal definition within the framework of Discrete Event Systems". *Reliability Engineering and System Safety*, Volume 114, pp.148-154.

Sørensen, J. N., Apostolakis, G. E., Kress, T. S., and Powers, D. A. (1999). "On the Role of Defense in Depth in Risk-Informed Regulation". In: Proceedings of the PSA '99. International topical meeting on probabilistic safety assessment, Washington, DC, August 22–26, 1999, American Nuclear Society, La Grange Park, Illinois. p. 408–413.

Sørensen, J. N., Apostolakis, G. E., and Powers, D. A. (2000). "On the role of safety culture in risk-informed regulation". In: Kondo S, Furuta K, editors. Psam 5: Probabilistic Safety Assessment and Management, Volumes 1–4, pp. 2205–2210.

Svedung, I., and Rasmussen, J. (2002). "Graphic representation of accident scenarios: mapping system structure and the causation of accidents". *Safety Science*, Volume 40, Issue 5, pp. 397–417.

Venkatasubramanian, V., Rengaswamy, R., Yin, K., and Kavuri, S. N. (2003a). “A review of process fault detection and diagnosis: Part I: Quantitative model-based methods”. *Computers & chemical engineering*, Volume *27*, Issue 3, pp. 293-311.

Venkatasubramanian, V., Rengaswamy, R., and Kavuri, S. N. (2003b). “A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies”. *Computers & Chemical Engineering*, Volume *27,* Issue 3, pp. 313-326.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., and Yin, K. (2003c). “A review of process fault detection and diagnosis: Part III: Process history based methods”. *Computers & Chemical Engineering*, Volume *27*, Issue 3, pp. 327-346.

Weick, K. E. & Sutcliffe, K. M. 2007. *Managing the unexpected : resilient performance in an age of uncertainty* (2nd ed.). San Francisco: Jossey-Bass.